Proceedings of the SDR 11 Technical Conference and Product Exposition,

IMPLEMENTATION OF PARALLEL LATTICE REDUCTION-AIDED MIMO DETECTOR USING GRAPHICS PROCESSING UNIT

Taehyun Kim (HY-SDR Research Center, Hanyang Univ., Seoul, South Korea; semoise6@dsplab.hanyang.ac.kr); Hyunwook Yang (HY-SDR Research Center, Hanyang Univ., Seoul, South Korea; ypk2pe@dsplab.hanyang.ac.kr); Chiyoung Ahn (HY-SDR Research Center, Hanyang Univ., Seoul, South Korea; ahncy@dsplab.hanyang.ac.kr); and Seungwon Choi* (corresponding author, HY-SDR Research Center, Hanyang Univ., Seoul, South Korea; choi@ieee.org)

ABSTRACT

Since H. Yao proposed Lattice Reduction (LR) algorithm for the MIMO detector, one can exploit the diversity gain provided by the LR method of which performance is quite comparable to that of Maximum Likelihood (ML) algorithm while the complexity is almost as simple as that of Zero Forcing (ZF) algorithm. In this paper, in order to further reduce the processing time of the LR-aided detector, Graphics Processing Unit (GPU) has been proposed as the main processor in such a way that the detection can be performed in parallel processing using multiple threads in GPU. We implemented a LR MIMO detector to verify that the LR-aided detector performs even faster than the ML detector for processing the WiMAX waveform, thus, real-time processing can be provided by the GPU-based LR MIMO detector.

1. INTRODUCTION

Recently, the spatial multiplexing MIMO technology has received an extremely keen interest as an efficient tool for providing high data rate. It is well known that the spatial multiplexing MIMO system can linearly increase the data throughput without additional frequency band or transmit power[1]. In MIMO system, we need a detector for estimating the transmit signal. Amongst various kinds of MIMO detectors, ML detector provides best performance because it can fully exploit the diversity gain. However, the complexity of ML algorithm increases exponentially as the number of antennas or modulation index increases[2]. In order to mitigate the complexity problem in ML, we can consider ZF or Minimum Mean Square Error (MMSE) detector[3],[4]. But, the performance of these linear algorithms is far worse than that of ML. LR algorithm has been newly researched for providing a performance comparable to that of ML with a lot less computational burden[5].

Amongst a few ways of implementing the LR algorithm, Lenstra-Lenstra-Lovasz (LLL) algorithm has been most widely used[6]. In this paper, we adopt complex LLL (CLLL) algorithm instead of normal LLL algorithm. Note that in CLLL the complex-valued channel matrix is used as it is while the channel matrix in LLL algorithm is converted into the corresponding real-valued one. It can also be observed that the entire complexity of CLLL algorithm is less than LLL algorithm because the channel matrix in LLL algorithm becomes 4 times larger than that of CLLL[7].

In this paper, we implement the MIMO detector adopting CLLL algorithm using GPU. Since the CLLL algorithm requires Gram-Schmidt orthogonalization and repeated procedures of column vector exchanging, its complexity is a lot higher than that of ZF, thus, a serial processor such as conventional Digital Signal Processor (DSP) does not seem to be a proper tool for implementation.

On the contrary, since GPU consists of multiple threads [8], we can save a lot of computation time if the LR algorithm can be processed at each of the threads in GPU simultaneously. We implemented the LR-aided detector using GPU and applied it to WiMAX MIMO system to verify a real-time applicability

The organization of this article is as follows. Section 2 explains the system structure and LR algorithm. Section 3 shows how to implement LR-aided detector on GPU. In Section 4, we provide a performance analysis of the implemented MIMO system in WiMAX signal environment. Section 5 concludes this paper.

2. SYSTEM MODEL

2.1. MIMO System Description

Figure 1 illustrates the MIMO system model used in this paper where n_T and n_R denote the number of transmit and receive antenna, respectively. Data bits are divided into n_T streams to form a transmit signal vector $\mathbf{x} = \left[x_1, \cdots x_{n_T} \right]^T$ through the demultiplexer and modulator shown in Figure 1.

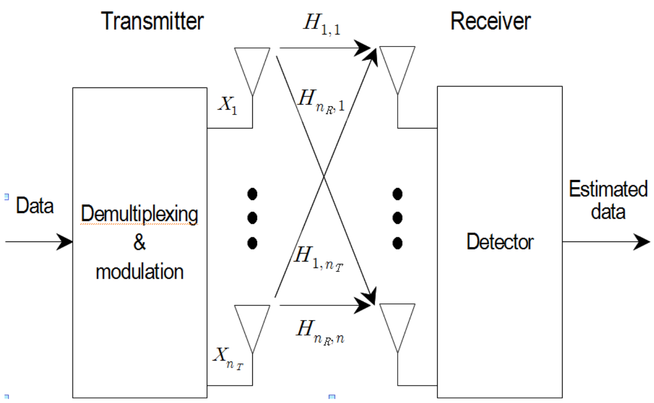

Figure 1. MIMO system model

In the receiver, the transmit data stream is retrieved using received signal vector $\mathbf{y}=\left[y_1,\ldots,y_{n_R}\right]^T$. Note that $[\ ]^T$ denotes transpose operator. The received signal vector can be represented as

$$\mathbf{y}=\mathbf{Hx}+\mathbf{n} \tag{1}$$

where $\mathbf{n}=\left[n_1,\ldots,n_{n_R}\right]^T$ is a Gaussian noise vector with i.i.d. (independent and identically distributed) elements and its variance σ_N^2, and $\mathbf{H}$ is $n_R \times n_T$ channel matrix.

2.2. Lattice Reduction-aided Detector

Lattice Reduction is to transform the channel matrix $\mathbf{H}$ into a nearly orthogonal matrix $\tilde{\mathbf{H}}$ using a unimodular matrix $\mathbf{T}$ consisting of integers. Note that $\det(\mathbf{T})$, determinant of $n_R \times n_T$ matrix $\mathbf{T}$, is +1 or -1.

In LR-aided detector, the transmitted signal $\mathbf{x}$ is estimated using $\tilde{\mathbf{H}}$ and $\mathbf{T}$. With $\tilde{\mathbf{H}}=\mathbf{HT}$ and $\mathbf{z}=\mathbf{T}^{-1}\mathbf{x}$, received signal vector can be written as

$$\mathbf{y}=\mathbf{Hx}+\mathbf{n}=\mathbf{HTT}^{-1}\mathbf{x}+\mathbf{n}=\tilde{\mathbf{H}}\mathbf{z}+\mathbf{n} \tag{2}$$

Note that the condition number of $\tilde{\mathbf{H}}$ is a lot less than that of $\mathbf{H}$ due to the better orthogonality, although $\mathbf{Hx}$ corresponds to the same point as $\tilde{\mathbf{H}}\mathbf{z}$ in lattice. Also, $\mathbf{z}$ is a constellation on the reduced lattice. Using $\tilde{\mathbf{H}}\mathbf{z}$, transmitted signal is detected through the following procedures.

$$\tilde{\mathbf{x}}=\mathbf{H}^{+}\mathbf{y}=\mathbf{x}+\mathbf{H}^{+}\mathbf{n} \tag{3}$$

$$\tilde{\mathbf{z}}=\mathbf{T}^{-1}\tilde{\mathbf{x}}=\tilde{\mathbf{H}}^{+}\mathbf{y}=\mathbf{z}+\tilde{\mathbf{H}}^{+}\mathbf{n} \tag{4}$$

where $\mathbf{H}^{+}=\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ denotes a channel matrix after the pseudo-inverse. Note that (3) represents a general ZF procedure while (4) is needed for LR-aided detector. Borrowing the quasi-orthogonality of $\tilde{\mathbf{H}}$, we can reduce the effect of noise enhancement when we multiply $\tilde{\mathbf{H}}^{+}$ to the noise compared to the other case when we multiply $\mathbf{H}^{+}$. Finally, the transmitted signal is estimated using $\mathbf{T}$ from $\tilde{\mathbf{z}}$, as $\tilde{\mathbf{x}}=\mathbf{T}\tilde{\mathbf{z}}$. Figure 2 illustrates a block diagram of LR-aided detector described above.

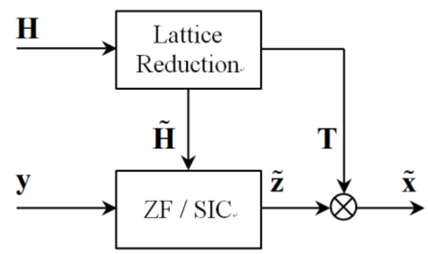

Figure 2. Block diagram of LR-aided detector

Table 1. Complex LLL Algorithm [9]

INPUT: $\mathbf{Q}$, $\mathbf{R}$ ($\mathbf{H}=\mathbf{QR}$)

OUTPUT: $\tilde{\mathbf{H}}=\tilde{\mathbf{Q}}\tilde{\mathbf{R}}$ which is LLL-reduced with parameter δ, $\mathbf{T}$ satisfying $\tilde{\mathbf{H}}=\mathbf{HT}=\mathbf{QRT}$

Initialization $\tilde{\mathbf{Q}}:=\mathbf{Q}$, $\tilde{\mathbf{R}}:=\mathbf{R}$, $\mathbf{T}:=\mathbf{I}_{n_T}$

$k:=2$

while $k \le n_T$

    for $j=k-1$ **to** 1 **step** -1

        $\mu=\tilde{R}_{j,k}/\tilde{R}_{j,j}$

        λ is the closest integer of μ

        $\tilde{\mathbf{R}}(1:j,k):=\tilde{\mathbf{R}}(1:j,k)-\lambda\tilde{\mathbf{R}}(1:j,j)$

        $\mathbf{T}(:,k):=\mathbf{T}(:,k)-\lambda\mathbf{T}(:,j)$

    end

    if $\delta\left|\tilde{R}_{k-1,k-1}\right|^2>\left|\tilde{R}_{k,k}\right|^2+\left|\tilde{R}_{k-1,k}\right|^2$

        swap columns $k-1$ and k in $\tilde{\mathbf{R}}$ and $\mathbf{T}$

        $\alpha=\dfrac{\tilde{R}_{k-1,k-1}}{\left\|\tilde{\mathbf{R}}(k-1:k,k-1)\right\|}$, $\beta=\dfrac{\tilde{R}_{k,k-1}}{\left\|\tilde{\mathbf{R}}(k-1:k,k-1)\right\|}$

        $\Theta=\begin{bmatrix}\bar{\alpha} & \beta \\ -\beta & \alpha\end{bmatrix}$

        $\tilde{\mathbf{R}}(k-1:k,k-1:n_T):=\Theta\tilde{\mathbf{R}}(k-1:k,k-1:n_T)$

        $\tilde{\mathbf{Q}}(:,k-1:k):=\tilde{\mathbf{Q}}(:,k-1:k)\Theta^H$

        $k:=\max(k-1,2)$

    else

        $k:=k+1$

    end

end

Conventional LLL algorithm adopts a channel matrix consisting of real-valued elements only, which causes the matrix size to be doubled. Computational load can be reduced tremendously if the complex-valued channel matrix $\mathbf{H}$ is adopted as it is. Table 1 shows the complex LLL (CLLL) [9]. Note that inputs $\mathbf{Q}$ and $\mathbf{R}$ have been obtained through the QR decomposition of $\mathbf{H}$ where $\mathbf{Q}$ is a unitary matrix and $\mathbf{R}$ is a upper triangular matrix, and $\bar{\alpha}$ is complex conjugate of α.

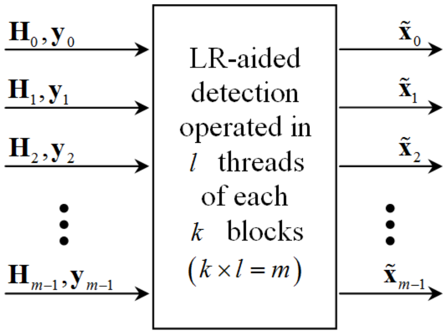

Figure 3. Structure of LR-aided detector implemented in GPU

3. IMPLEMENTATION

In this section, an LR-aided detector is implemented using GPU for a speed-up operation. GPU having a Single Instruction Multiple Data (SIMD) structure is very advantageous in parallel processing because a single instruction can process multiple data. Inside GPU, there are a number of threads at each block where multiple blocks form an entire structure such that plenty of threads perform each instruction with parallel processing, which eventually reduce the operation time tremendously.

Figure 3 illustrates the structure of implemented LR-aided detector. Once m data are input to the corresponding m threads after parallelizing received signals, LR-aided detector processes the input data in parallel and transfers decoded data to the next step at once.

In order to confirm the feasibility of proposed implementation in practical situations, the LR-aided detector is ported on the MIMO detector blocks in WiMAX system that is implemented using GPU. Since the implemented WiMAX system adopts 2X2 MIMO with 16QAM modulation, the total number of symbols at each frame of MIMO decoder block is 17,280. It particularly means that experimental tests have been performed with k blocks and l threads such that the total number, $m = k \times l$, is 17,280.

When GPU is adopted in the implementation, the number of blocks and threads becomes a major factor of operation time. Due to the internal structure, the number of threads is to be a multiple of 32 and the operation speed of GPU becomes faster as the number of threads per block increases[10]. Figure 4 illustrates the operation time of MIMO decoder according to the number of blocks and threads. As shown in the figure, the operation time is reduced as the number of threads, l, becomes a multiple of 32 and as l increases. In our implementation, the number of blocks and that of threads have been set up to be 36 and 480, respectively, such that the MIMO detector operates within a minimum time. Note that the maximum number of threads is 512.

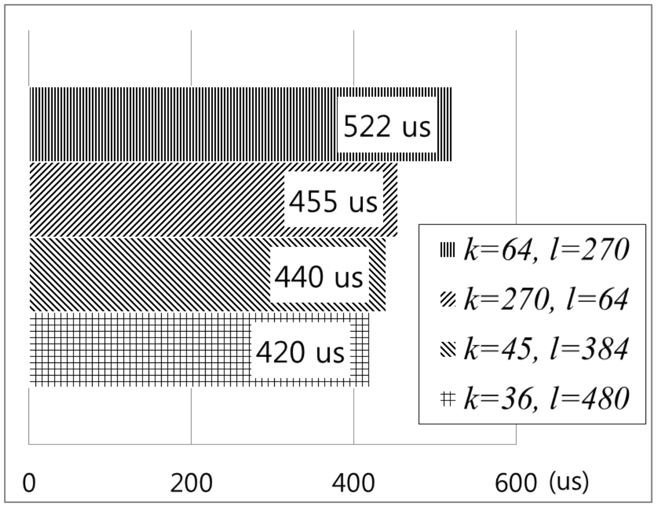

Figure 4. Operation time of MIMO decoder according to the number of blocks and threads

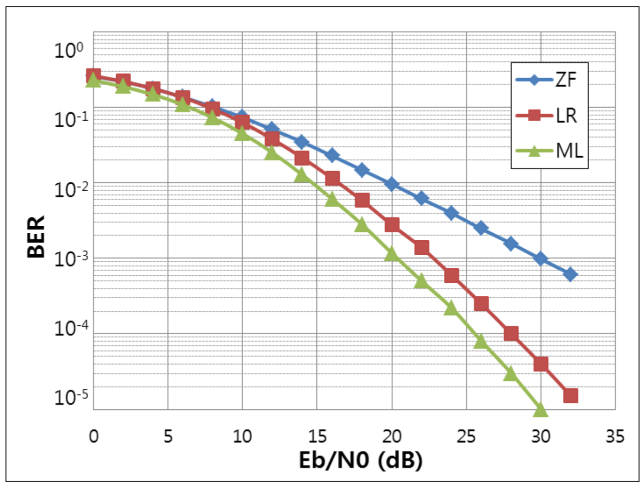

Figure 5. BER performance of ML, ZF and LR ($n_T = n_R = 2$, 16QAM, uncoded)

4. PERFORMANCE ANALYSIS

In this section, we show a performance of LR-aided MIMO detector implemented with GPU. In our analysis which has been performed in 2X2 MIMO system adopting 16QAM, we compare the proposed LR-aided MIMO detector with ML and ZF detector in WiMAX signal environment. Figure 5 illustrates the performance of the 3 detectors showing that the proposed LR-aided detector exhibits a comparable performance to the ML detector while ZF is far behind.

4.1. Operation Time

In order to measure the operation time taken in GPU, we used the profiler provided by GPU provider[10]. Table 2 shows the operation time needed for the LR-aided detector to process a single symbol and frame. As shown in the table,

Table 2. Operation time of LR-aided detector needed for processing 1 symbol and 1 frame

1 symbol	1 frame (17,280 symbols)
0.07ms	0.42ms

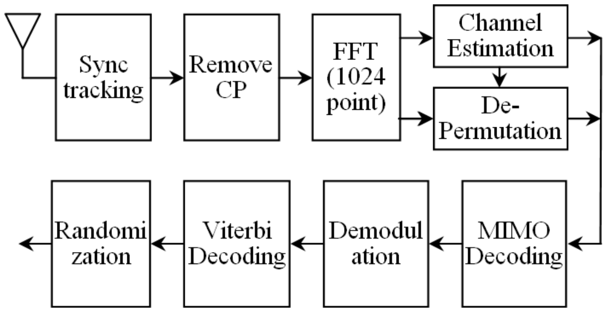

Figure 6. Block diagram of WiMAX receiver system

it takes about 0.07ms for processing 1 symbol. It is noteworthy that it would take 17,280 times 0.07ms for processing a single frame, if a serial processor instead of GPU has been used for the implementation. Due to the parallel processing provided by GPU, it takes only 6 times, i.e., 0.42ms, for processing a frame which contains 17,280 symbols.

Figure 6 illustrates a block diagram of WiMAX receiver system. We measured the operation time using the WiMAX system shown in Figure 6 after adopting the proposed LR-aided detector for the MIMO decoding block.

Table 3 shows the operation time taken at the receiver of Figure 6[11] as well as the specific time required for the MIMO detector to process 1 frame data when each of LR-aided, ML, and ZF algorithm is adopted. As shown in the table, the proposed LR-aided detector is about 0.4ms slower than ZF, but, faster than ML by about 9 times.

As is well known, WiMAX is a TDD(Time Division Duplex) system with 3ms and 2ms being assigned for downlink and uplink, respectively. In our implementation, we realized the downlink only, meaning that the total processing time should not exceed 3ms. From Table 3, it can be observed that the proposed LR-aided detector, consuming about 2.75ms in total, can be applied for real WiMAX adopting 2X2 MIMO with 16QAM modulation while ML is not practically applicable because the total processing time far exceeds 3ms.

5. CONCLUSION

We have implemented 2X2 MIMO WiMAX system with 16QAM using GPU. CLLL algorithm instead of conventional LLL algorithm has been adopted for implementing LR-aided detector for the MIMO WiMAX system. CLLL algorithm together with the other modem

Table 3. Comparison of operation time of ML, LR and ZF algorithm on WiMAX system

Algorithm	ML	LR	ZF
MIMO detector	3.84ms	0.42ms	0.04ms
all blocks except detector in the receiver	2.33ms		
Total	6.17ms	2.75ms	2.37ms

procedures of WiMAX have been coded on GPU for parallel processing using multiple threads provided in the GPU. Exploiting the multiple blocks and threads inside the GPU, we demonstrated that the operation time required for implementing the LR-aided detector is short enough for real-time processing of WiMAX (2X2 MIMO with 16QAM). We also showed that the LR-aided detector outperforms ZF detector. The performance of the LR-aided detector is quite comparable to that of ML detector while the processing time of the former is less than that of the latter by nearly 9 times. From the results obtained from our implementation, we conclude that LR-aided detector implemented on GPU can be widely applied to future communication systems which need extremely fast processing time.

6. REFERENCES

[1] S. Haykin and M. Moher, *Modern Wireless Communications*, Pearson Prentice Hall, New Jersey, 2005.

[2] S.Verdu, *Multiuser Detection*. Cambridge, U.K: Cambridge Univ. Press, 1998.

[3] J. Proakis, *Digital communications*, 4th ed., New York, NY, McGraw- Hill, 2000

[4] D. Wübben, R. Böhnke, V. Kühn, and K. Kammeyer, "Near-Maximum-Likelihood Detection of MIMO Systems using MMSE-based Lattice-Reduction," *2004 IEEE International Conference on Communications*, Vol. 2, pp. 798-802, 2004.

[5] H. Yao and W. Wornell, "Lattice-Reduction-Aided Detectors for MIMO Communication Systems," *Global Telecommunication Conference*, Vol. 1, pp. 424-428, 2002.

[6] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring Polynomials with Rational Coefficients," *Math. Ann*, Vol. 261, pp. 515-534, 1982.

[7] Y. H. Gan and W. H. Mow, “Complex Lattice Reduction Algorithms for Low-Complexity MIMO Detection,” *IEEE Globecom 2005*, pp. 2953-2957, 2005

[8] J. Kim, S. Hyeon, and S. Choi, “Implementation of an SDR System Using Graphics Processing Unit,” *Communications Magazine of the IEEE*, Vol. 48, 2010.

[9] H. Vetter, V. Ponnampalam, M. Sandell, and P. A. Hoeher, “Fixed Complexity LLL Algorithm,” *IEEE Transactions on Signal Processing*, Vol. 57, NO. 4, April 2009

[10] NVIDIA Corporation, CUDA C Programming Guide, 2010.

[11] J. Ju, C. Ahn, J. Kim, S. Hyeon, and S. Choi, “Implementation of an SDR System Using GPU and Its Application to 2x2 MIMO WiMAX System,” 2010 Wireless Innovation Forum, pp.341-345, Dec 2010